

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-202889

(43)Date of publication of application : 30.07.1999

(51)Int.Cl. G10L 3/00
G10L 3/00
G06F 3/16
G09B 7/02
G09B 19/04

(21)Application number : 10-253449

(71)Applicant : INTERNATL BUSINESS MACH CORP <IBM>

(22)Date of filing : 08.09.1998

(72)Inventor : MINEMATSU AYAKO

(30)Priority

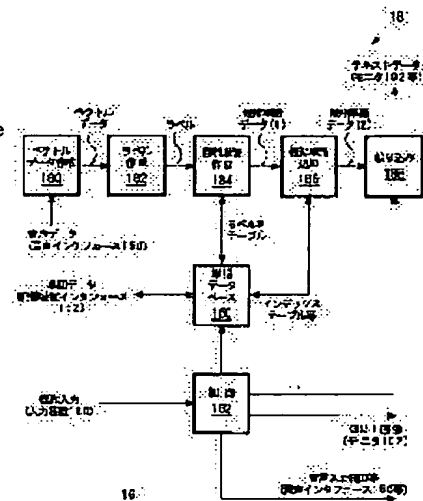
Priority number : 09315519 Priority date : 17.11.1997 Priority country : JP

(54) SPEECH DISCRIMINATING DEVICE, AND DEVICE AND METHOD FOR PRONUNCIATION CORRECTION

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a device and method for speech discrimination which accurately discriminates an English word from English pronunciation by, for example, a nonnative.

SOLUTION: A vector data process part 160 and a label generation part 182 process and convert speech data of one sentence of English words that a Japanese has pronounced into a label sequence. A candidate word generation part 184 makes the label sequence of one sentence correspond to a 1st candidate word consisting of not less than one English word. A similar word addition part 186 retrieves an English word similar to the pronunciation of the 1st candidate word, e.g. a word 'lead' similar to a 1st candidate word 'read' (In general, it is difficult for a Japanese to distinctively pronounce 'l' and 'r'.) by using a word data base 160 and adds the obtained similar word to the 1st candidate word to obtain a 2nd candidate word. A narrowing-down part 188 selects one of 2nd candidate words as a final discrimination result according to user's operation, connects selected English words, and outputs them as English text data.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japan Patent Office

BEST AVAILABLE COPY

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-202889

(43) 公開日 平成11年(1999) 7月30日

(51) Int.Cl.⁸
 G 1 0 L 3/00
 G 0 6 F 3/16
 G 0 9 B 7/02
 19/04

識別記号
 5 6 1
 5 5 1
 3 2 0

F I
 G 1 0 L 3/00
 G 0 6 F 3/16
 G 0 9 B 7/02
 19/04

審査請求 未請求 請求項の数17 OL (全 22 頁)

(21) 出願番号 特願平10-253449
 (22) 出願日 平成10年(1998) 9月8日
 (31) 優先権主張番号 特願平9-315519
 (32) 優先日 平9 (1997) 11月17日
 (33) 優先権主張国 日本 (J P)

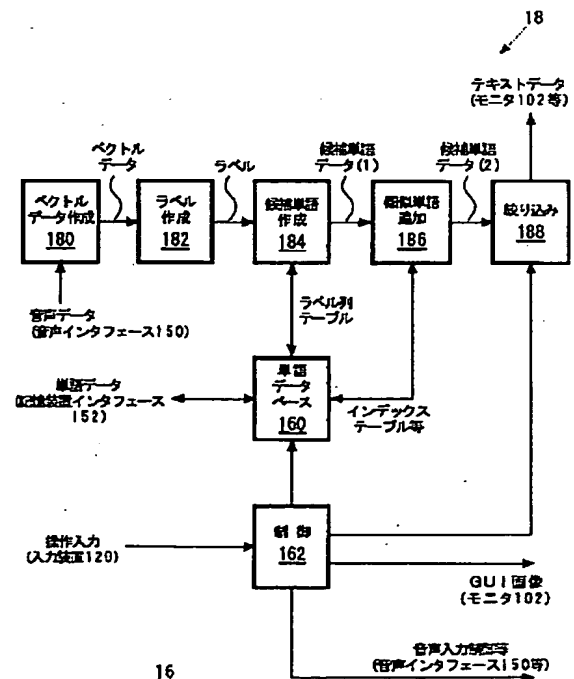
(71) 出願人 390009531
 インターナショナル・ビジネス・マシーン
 ズ・コーポレーション
 INTERNATIONAL BUSIN
 ESS MASCHINES CORPO
 RATION
 アメリカ合衆国10504、ニューヨーク州
 アーモンク (番地なし)
 (72) 発明者 峰 松 彩 子
 神奈川県大和市下鶴間1623番地14 日本ア
 イ・ビー・エム株式会社 大和事業所内
 (74) 代理人 弁理士 坂口 博 (外1名)

(54) 【発明の名称】 音声識別装置、発音矯正装置およびこれらの方法

(57) 【要約】

【課題】例えば、ノン・ネイティブの英語の発音から、正確に英単語を識別する音声識別装置およびその方法を提供することを目的とする。

【解決手段】ベクトルデータ生成部160およびラベル作成部182は、日本人が発音した英語の1センテンス分の音声データを処理し、ラベル列に変換する。候補単語作成部184は、1センテンス分のラベル列を、それぞれ1つ以上の英単語から構成される第1の候補単語に対応付ける。類似単語追加部186は、単語データベース部160を用いて、第1の候補単語の発音に類似する英単語、例えば、第1の候補単語readに対する類似単語lead (日本人は一般にlとrとを区別して発音しづらい)を検索し、得られた類似単語を第1の候補単語に付加して第2の候補単語とする。絞り込み部188は、ユーザの操作に応じて第2の候補単語のいずれかを最終的な識別結果として選択し、選択した英単語を接続して英文のテキストデータとして出力する。



1

【特許請求の範囲】

【請求項1】話し声に含まれる1つ以上の単語を示す音声データから、前記単語それぞれを識別する音声識別装置であって、

1つ以上の前記単語の音声データそれぞれに、これら1つ以上の音声データそれぞれを識別して得られる1つ以上の前記単語の組み合わせの候補（候補単語）の1組以上を対応付ける候補単語対応付け手段と、

1つ以上の前記単語の音声データそれぞれに対応付けた前記候補単語それぞれに、前記候補単語それぞれの発音に対応しうる1つ以上の前記単語の組み合わせ（類似単語）の0組以上を対応付ける類似単語対応付け手段と、

1つ以上の前記単語の音声データそれぞれに対応付けた前記候補単語、および、前記候補単語それぞれに対応付けた前記類似単語のいずれかを選択し、前記単語の音声データそれぞれの識別結果とする音声データ識別手段とを有する音声識別装置。

【請求項2】前記音声データは、所定の言語の話し声に含まれる1つ以上の単語を示し、

前記候補単語対応付け手段は、1つ以上の前記単語の音声データそれぞれに、これら1つ以上の音声データそれぞれを識別して得られる前記所定の言語の候補単語の1組以上を対応付け、

前記類似単語対応付け手段は、1つ以上の前記単語の音声データそれぞれに対応付けた前記候補単語それぞれに、前記候補単語それぞれの発音に対応しうる前記所定の言語の類似単語の0組以上を対応付け、

音声データ識別手段は、1つ以上の前記単語の音声データそれぞれに対応付けた前記候補単語、および、前記候補単語それぞれに対応付けた前記類似単語のいずれかを選択し、1つ以上の前記単語の音声データそれぞれの識別結果とする請求項1に記載の音声識別装置。

【請求項3】前記所定の言語の話し声の発音は、前記所定の言語以外の所定の言語を主に話す話者によってなされ、

前記話者によって前記所定の言語の単語の1つ以上それぞれが発音された場合に、前記所定の言語の話し声に含まれる単語の音声データの1つ以上それぞれに対応しうる前記所定の言語の単語の0組以上を、予め前記所定の言語の単語の1つ以上それぞれと対応付けて、前記所定の言語の単語の1つ以上それぞれの前記類似単語として記憶する類似単語記憶手段を有し、

前記類似単語対応付け手段は、予め前記所定の言語の単語の1つ以上それぞれと対応付けて記憶した前記類似単語の0組以上を、前記候補単語それぞれに対応付ける請求項2に記載の音声識別装置。

【請求項4】前記候補単語対応付け手段は、音声データに対応付けた前記候補単語それぞれの確からしさを示す確率データを、音声データに対応付けた前記候補単語それぞれに付し、

2

前記音声データ識別手段は、前記確率データの値が所定の範囲をとる前記候補単語のみを選択し、前記単語の音声データの識別結果とする請求項3に記載の音声識別装置。

【請求項5】前記候補単語対応付け手段は、前記類似単語それぞれに対応する発音の誤りを示す誤り情報を、音声データに対応付けた前記候補単語それぞれに付す請求項3に記載の音声識別装置。

【請求項6】所定の言語以外を主に話す話者による前記所定の言語の音声に含まれる前記所定の言語の1つ以上の単語を示す音声データから、前記所定の言語の単語の1つ以上それぞれを識別する音声識別装置であって、

前記所定の言語の単語の音声データの1つ以上それぞれに、これらの音声データの1つ以上それぞれを識別して得られる前記所定の言語の単語、および、前記話者が話した可能性がある前記所定の言語の単語の1つ以上またはこれらのいずれかに対応付ける単語対応付け手段と、

1つ以上の前記単語の音声データそれぞれに対応付けた単語のいずれかを選択し、1つ以上の前記単語の音声データそれぞれの識別結果とする音声データ識別手段とを有する音声識別装置。

【請求項7】話し声に含まれる1つ以上の単語を示す音声データから、前記単語それぞれを識別する音声識別方法であって、

1つ以上の前記単語の音声データそれぞれに、これらの音声データの1つ以上それぞれを識別して得られる1つ以上の前記単語の組み合わせの候補（候補単語）の1組以上を対応付けるステップと、

1つ以上の前記単語の音声データそれぞれに対応付けた前記候補単語それぞれに、前記候補単語それぞれの発音に対応しうる1つ以上の前記単語の組み合わせ（類似単語）の0組以上を対応付けるステップと、

1つ以上の前記単語の音声データそれぞれに対応付けた前記候補単語、および、前記候補単語それぞれに対応付けた前記類似単語のいずれかを選択し、前記単語の音声データそれぞれの識別結果とするステップとを含む音声識別方法。

【請求項8】所定の言語以外を主に話す話者による前記所定の言語の音声に含まれる前記所定の言語の1つ以上の単語を示す音声データから、前記所定の言語の単語の1つ以上それぞれを識別する音声識別方法であって、

前記所定の言語の単語の音声データの1つ以上それぞれに、これらの音声データの1つ以上それぞれを識別して得られる前記所定の言語の単語、および、前記話者が話した可能性がある前記所定の言語の単語の1つ以上またはこれらのいずれかに対応付けるステップと、

1つ以上の前記単語の音声データそれぞれに対応付けた単語のいずれかを選択し、1つ以上の前記単語の音声データそれぞれの識別結果とするステップとを含む音声識別方法。

【請求項9】話し声に含まれる1つ以上の単語を示す音声データから、前記単語の1つ以上それぞれを識別する音声識別装置において、

1つ以上の前記単語の音声データそれぞれに、これらの音声データの1つ以上それぞれを識別して得られる1つ以上の前記単語の組み合わせの候補（候補単語）の1組以上を対応付ける候補単語対応付けステップと、

1つ以上の前記単語の音声データそれぞれに対応付けた前記候補単語それぞれに、前記候補単語それぞれの発音に対応しうる1つ以上の前記単語の組み合わせ（類似単語）の0組以上を対応付ける類似単語対応付けステップと、

1つ以上の前記単語の音声データそれぞれに対応付けた前記候補単語、および、前記候補単語それぞれに対応付けた前記類似単語のいずれかを選択し、前記単語の音声データそれぞれの識別結果とする音声データ識別ステップとをコンピュータに実行させるプログラムを記録したコンピュータにより読み取り可能な記録媒体。

【請求項10】前記音声データは、所定の言語の話し声に含まれる1つ以上の単語を示し、

前記候補単語対応付けステップにおいて、1つ以上の前記単語の音声データそれぞれに、これらの音声データの1つ以上それぞれを識別して得られる前記所定の言語の候補単語の1組以上を対応付け、

前記類似単語対応付けステップにおいて、1つ以上の前記単語の音声データそれぞれに対応付けた前記候補単語それぞれに、前記候補単語それぞれの発音に対応しうる前記所定の言語の類似単語の0組以上を対応付け、音声データ識別ステップにおいて、1つ以上の前記単語の音声データそれぞれに対応付けた前記候補単語、および、前記候補単語それぞれに対応付けた前記類似単語のいずれかを選択し、前記単語の音声データそれぞれの識別結果とする処理をコンピュータに実行させるプログラムを記録したコンピュータにより読み取り可能な請求項9に記載の記録媒体。

【請求項11】前記所定の言語の話し声は、前記所定の言語以外の所定の言語を主に話す話者によってなされ、前記話者によって前記所定の言語の単語の1つ以上それぞれが発音された場合に、前記所定の言語の話し声に含まれる単語の音声データの1つ以上それぞれに対応しうる前記所定の言語の単語の0組以上を、予め前記所定の言語の単語の1つ以上それぞれと対応付けて、前記所定の言語の単語の1つ以上それぞれの前記類似単語として記憶し、

前記類似単語対応付けステップにおいて、予め前記所定の言語の単語の1つ以上それぞれと対応付けて記憶した前記類似単語の0組以上を、前記候補単語それぞれに対応付ける処理をコンピュータに実行させるプログラムを記録したコンピュータにより読み取り可能な請求項10に記載の記録媒体。

【請求項12】前記候補単語対応付けステップにおいて、音声データに対応付けた前記候補単語それぞれの確からしさを示す確率データを、音声データに対応付けた前記候補単語それぞれに付し、

前記音声データ識別ステップにおいて、前記確率データの値が所定の範囲をとる前記候補単語のみを選択し、前記単語の音声データの識別結果とする処理をコンピュータに実行させるプログラムを記録したコンピュータにより読み取り可能な請求項11に記載の記録媒体。

【請求項13】前記候補単語対応付けステップにおいて、前記候補単語それぞれに対応する発音の誤りを示す誤り情報を、音声データに対応付けた前記類似単語それぞれに付す処理をコンピュータに実行させるプログラムを記録したコンピュータにより読み取り可能な請求項11に記載の記録媒体。

【請求項14】所定の言語以外を主に話す話者による前記所定の言語の音声に含まれる前記所定の言語の1つ以上の単語を示す音声データから、前記所定の言語の単語それぞれを識別する音声識別装置において、

前記所定の言語の単語の音声データの1つ以上それぞれに、これらの音声データの1つ以上それぞれを識別して得られる前記所定の言語の単語の1つ以上、および、前記話者が話した可能性がある前記所定の言語の単語の1つ以上またはこれらのいずれかを対応付ける単語対応付けステップと、

1つ以上の前記単語の音声データそれぞれに対応付けた単語のいずれかを選択し、1つ以上の前記単語の音声データそれぞれの識別結果とする音声データ識別ステップとをコンピュータに実行させるプログラムを記録したコンピュータにより読み取り可能な記録媒体。

【請求項15】単語を示す音声データを識別して得られる単語の候補（候補単語）を1個以上、対応付ける候補単語対応付け手段と、

音声データに対応付けた前記候補単語それぞれに、前記候補単語それぞれの発音に対応しうる単語（類似単語）を0個以上、対応付ける類似単語対応付け手段と、前記音声データが示す単語と、この音声データに対応付けられた前記候補単語それぞれに対応付けられた前記類似単語とが一致する場合に、前記音声データが示す単語と同じ前記類似単語に対応し、前記音声データが示す単語の発音を矯正する発音矯正データを出力する発音矯正データ出力手段とを有する発音矯正装置。

【請求項16】単語を示す音声データを識別して得られる単語の候補（候補単語）を1個以上、対応付け、音声データに対応付けた前記候補単語それぞれに、前記候補単語それぞれの発音に対応しうる単語（類似単語）を0個以上、対応付け、

前記音声データが示す単語と、この音声データに対応付けられた前記候補単語それぞれに対応付けられた前記類似単語とが一致する場合に、前記音声データが示す単語

と同じ前記類似単語に対応し、前記音声データが示す単語の発音を矯正する発音矯正データを出力する発音矯正方法。

【請求項17】単語を示す音声データを識別して得られる単語の候補（候補単語）を1個以上、対応付ける候補単語対応付けステップと、
音声データに対応付けた前記候補単語それぞれに、前記候補単語それぞれの発音に対応しうる単語（類似単語）を0個以上、対応付ける類似単語対応付けステップと、前記音声データが示す単語と、この音声データに対応付けられた前記候補単語それぞれに対応付けられた前記類似単語とが一致する場合に、前記音声データが示す単語と同じ前記類似単語に対応し、前記音声データが示す単語の発音を矯正する発音矯正データを出力する発音矯正データ出力ステップと処理をコンピュータに実行させるプログラムを記録したコンピュータにより読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、特定の言語を母国語とする話者による他の特定の言語の話し声に含まれる後者の言語の単語それぞれを識別する音声識別装置およびその方法、例えば、日本人が話す英語の音声を識別し、音声に含まれる英単語の列を示すデータ（テキストデータ）として出力する音声識別装置およびその方法に関する。

【0002】また、本発明は、上記音声識別装置およびその方法の処理過程において得られるデータ（候補単語データ）を利用して、正しい発音を話者に教え、発音を矯正する発音矯正装置およびその方法に関する。

【0003】

【従来の技術】不特定話者が話す音声に含まれる単語それぞれを識別し、テキストデータとして出力する音声識別装置が用いられている。例えば、特開平06-12483号公報、特開平08-50493号公報および特開平09-22297（文献1〜3）等は、このような音声識別方法を開示する。

【0004】例えば、これら従来の音声識別方法を用い、英語の音声から英単語を識別する英語用音声識別装置により、日本人が発音した英語音声から英文のテキストデータを作成すると、識別率が低下してしまう。これは、英語には、日本語に存在しない音（th等）、あるいは、日本語においては区別しにくい音（l,r等）等があり、日本人は、一般に、このような英語の音を正確に発音できないので、英語用音声識別装置が不正確な発音をそのまま単語に置き換えてしまうためである。例えば、日本人が英語で“rice”と発音したつもりであっても、英語用の音声識別装置は、この音声を“lice”あるいは“lou se”と識別してしまうことがある。

【0005】このような不具合は、上に示した例とは逆

に、英語を母国語とするアメリカ人が、日本語の音声から日本語のテキストを作成する音声識別装置を用いる場合、あるいは、イギリス式の英語を母国語とする英国人が、アメリカ式英語用に調整された音声識別装置を用いる場合、あるいは、特定の人が何らかの理由で正確な発音をしにくくなった場合等、様々な場合に生じうる。しかしながら、上記各文献に開示された音声識別方法のいずれも、このような不具合を解決しえない。

【0006】ここで、話者の英語の発音が上達し、ネイティブスピーカーに近づけば、音声識別装置による識別率が向上することは言うまでもなく、しかも、話者にとって英会話が上達することは好ましいことである。

【0007】例えば、特開平4-54956号公報は、話者の英語音声を識別し、識別した英語音声を話者に確認させる学習装置を開示する（文献4）。また、例えば、特開昭60-123884号公報は、音声合成LSIを用い、話者に学習しようとする音声を聞かせる英語学習機を開示する（文献5）。その他、特公昭44-7162号公報、特開平7-117807号公報、特開昭61-18068号公報、特開平8-27588号公報、特開昭62-111278号公報、特開昭62-299985号公報、特開平3-75869号公報、特公平6-27971号公報、特公平8-12535号公報および特開平3-226785号公報等、数多くの文献に、外国語の発音を学習するための学習装置が開示されている（文献6〜14）。

【0008】しかしながら、これらの文献に開示された学習装置を用いても、提示された発音と自分の発音とを比較しなければならなかったり、自分の発音のどこが誤っているかが分からなかったりして、話者が十分な学習効果を必ずしも得ることはできない。

【0009】

【発明が解決しようとする課題】本発明は、上述した従来技術の問題点に鑑みてなされたものであり、所定の言語を母国語としない話者（ノン・ネイティブ）による所定の言語の話し声に含まれる単語それぞれを識別し、話者が意図する所定の言語の単語に置換して、正確なテキストデータを作成することができる音声識別装置およびその方法を提供することを目的とする。

【0010】また、本発明は、話されている地域が異なる等のために、同一の言語の発音が変化したような場合であっても、いずれの地域の話者による話し声でも、話者が意図する単語に変換して、正確なテキストデータを作成することができる音声識別装置およびその方法を提供することを目的とする。また、本発明は、発音の個人差を補って、常に高い識別率を保つことができる音声識別装置およびその方法を提供することを目的とする。

【0011】さらに、本発明は、上記音声識別装置およびその方法の処理の過程で得られるデータを利用して話者の発音の問題点を指摘し、話者にネイティブスピーカ

の発音を学習させ、話者の発音を矯正する発音矯正装置およびその方法を提供することを目的とする。また、本発明は、話者の発音と正確な発音とを自動的に比較して誤りを指摘することができ、さらに、話者が発音をどのように矯正すべきかを示す詳細な情報を提示し、その発音を矯正することができる発音矯正装置およびその方法を提供することを目的とする。

【0012】

【課題を達成するための手段】【第1の音声識別装置】

上記目的を達成するために、本発明にかかる第1の音声識別装置は、話し声に含まれる1つ以上の単語を示す音声データから、前記単語それぞれを識別する音声識別装置であって、1つ以上の前記単語の音声データそれぞれに、これらの音声データの1つ以上それぞれを識別して得られる1つ以上の前記単語の組み合わせの候補（候補単語）の1組以上を対応付ける候補単語対応付け手段と、1つ以上の前記単語の音声データそれぞれに対応付けた前記候補単語それぞれに、前記候補単語それぞれの発音に対応しうる1つ以上の前記単語の組み合わせ（類似単語）の0組以上を対応付ける類似単語対応付け手段と、1つ以上の前記単語の音声データそれぞれに対応付けた前記候補単語、および、前記候補単語それぞれに対応付けた前記類似単語のいずれかを選択し、前記単語の音声データそれぞれの識別結果とする音声データ識別手段とを有する。

【0013】好適には、前記音声データは、所定の言語の話し声に含まれる1つ以上の単語を示し、前記候補単語対応付け手段は、1つ以上の前記単語の音声データそれぞれに、これらの音声データの1つ以上それぞれを識別して得られる前記所定の言語の候補単語の1組以上を対応付け、前記類似単語対応付け手段は、1つ以上の前記単語の音声データそれぞれに対応付けた前記候補単語それぞれに、前記候補単語それぞれの発音に対応しうる前記所定の言語の類似単語の0組以上を対応付け、音声データ識別手段は、1つ以上の前記単語の音声データそれぞれに対応付けた前記候補単語、および、前記候補単語それぞれに対応付けた前記類似単語のいずれかを選択し、1つ以上の前記単語の音声データそれぞれの識別結果とする。

【0014】好適には、前記所定の言語の話し声の発音は、前記所定の言語以外の所定の言語を主に話す話者によってなされ、前記話者によって前記所定の言語の単語の1つ以上それぞれが発音された場合に、前記所定の言語の話し声に含まれる単語の音声データの1つ以上それぞれに対応しうる前記所定の言語の単語の0組以上を、予め前記所定の言語の単語の1つ以上それぞれと対応付けて、前記所定の言語の単語の1つ以上それぞれの前記類似単語として記憶する類似単語記憶手段を有し、前記類似単語対応付け手段は、予め前記所定の言語の単語の1つ以上それぞれと対応付けて記憶した前記類似単語の

0組以上を、前記候補単語それぞれに対応付ける。

【0015】好適には、前記候補単語対応付け手段は、音声データに対応付けた前記候補単語それぞれの確からしさを示す確率データを、音声データに対応付けた前記候補単語それぞれに付し、前記音声データ識別手段は、前記確率データの値が所定の範囲をとる前記候補単語のみを選択し、前記単語の音声データの識別結果とする。

【0016】好適には、前記候補単語対応付け手段は、前記類似単語それぞれに対応する発音の誤りを示す誤り情報を、音声データに対応付けた前記候補単語それぞれに付す。

【0017】【音声識別装置の作用】本発明にかかる音声識別装置は、特定の言語（以下本項目において英語を例示する）以外の言語（同様に日本語を例示する）を母国語とし、主に話す話者（同様に日本人を例示する）が発音した英語の話し声（音声）に含まれる単語を識別し、英単語に置き換えてテキストデータを作成する。

【0018】本発明にかかる音声識別装置において、マイク等から入力され、デジタルデータに変換された日本人による英語音声（音声データ）は、例えば、音の特徴（音の高さ、強さ、イントネーション等）ごとに量子化したベクトルデータに変換され、さらに、ラベルと呼ばれ、発音記号に類似する音のデータに変換されて、候補単語対応付け手段に対して出力される。

【0019】【候補単語対応付け手段】候補単語対応付け手段は、ラベルに変換された音声データを1つの単語、あるいは、一連の複数の単語ごとに処理し、音声データと、音声データの識別結果の候補として、単独（1個）の英単語、あるいは、複数の英単語の組み合わせ（これらを総称して候補単語と記す）とを対応付ける。

【0020】【類似単語記憶手段】類似単語記憶手段は、正確な英語の発音とは異なるが、日本人が英語を発音した場合に、音声データと対応しうる単独の英単語、あるいは、複数の英単語の組み合わせ（これらを総称して類似単語と記す）と、候補単語となりうる単独の英単語あるいは複数の英単語の組み合わせとを予め対応付けた、例えば、検索用辞書データを記憶する。一例を挙げると、この辞書データにおいては、日本人による不正確な英語の発音に対応するために、候補単語となりうる単独の英単語“lead”に、類似単語“read”（日本人が区別しにくいl,rを考慮したもの、一般に日本人は“r”の発音が不得手である）が対応付けられる。なお、英単語に対して類似単語が存在しない場合があるので、このような場合には、この辞書において、英単語に類似単語は対応付けられない。

【0021】【類似単語対応付け手段】類似単語対応付け手段は、類似単語記憶手段が記憶した辞書データを検索し、候補単語に対応付けられた類似単語を読み出して、候補単語に対する類似単語の対応付けを行なう。上記例においては、日本人が発音した英単語“read”に対応

する音声データに、英単語“lead”と、類似単語“read”とが対応付けられる。

【0022】[音声データ識別手段] 音声データ識別手段は、例えば、それまでに識別された英単語列に対する構文解析処理に基づいて、あるいは、ユーザの選択操作に応じて、音声データに対応付けられた候補単語および類似単語のいずれかを選択し、識別結果とする。

【0023】本発明にかかる音声識別装置の各構成要素は、次々に入力される音声データに対して、ここまでに述べた処理を順次、行ない、音声データに含まれる英単語を識別し、識別した英単語を接続したテキストデータを作成する。

【0024】なお、ここまでの説明では、日本人による英語の音声为例示したが、類似単語記憶手段が、アメリカ式の英語の発音とは異なるが、イギリス式の英語の発音である場合に、音声データと対応しうる類似単語を、候補単語に対応付けた辞書データを記憶する等の変更を加えることにより、本発明にかかる音声識別装置は、イギリス式の発音による英語音声、および、アメリカ式の発音による英語音声の両者を識別してテキストデータを生成することができる。

【0025】[所定の言語の範囲] このように、上記「所定の言語」の範囲は、候補単語対応付け手段が、充分な識別率で、音声データと単語とを対応付けられる範囲として定義される。従って、例えば、通常、同一とされる言語であっても、地域的な隔離により発音が相互に異なってしまう、いずれかに適するように調整された候補単語対応付け手段のみでは充分な識別率を得られない複数の方言(例えば、アメリカ、イギリス、オーストラリアおよび南アフリカ等各国の英語、スペインおよび中南米等各国のスペイン語)は、上記「所定の言語」の同一範囲には含まれない。さらに、何らかの理由により、特定の人の発音が不明りょうになり、この人の母国語(主に話す言葉)に適するように調整された候補単語対応付け手段のみでは充分な識別率が得られなくなった場合も同様である。

【0026】[第2の音声識別装置] また、本発明にかかる第2の音声識別装置は、所定の言語以外を主に話す話者による前記所定の言語の音声に含まれる前記所定の言語の1つ以上の単語を示す音声データから、前記所定の言語の単語の1つ以上それぞれを識別する音声識別装置であって、前記所定の言語の単語の音声データの1つ以上それぞれに、これらの音声データの1つ以上それぞれを識別して得られる前記所定の言語の単語、および、前記話者が話した可能性がある前記所定の言語の単語の1つ以上またはこれらのいずれかを対応付ける単語対応付け手段と、1つ以上の前記単語の音声データそれぞれに対応付けた単語のいずれかを選択し、1つ以上の前記単語の音声データそれぞれの識別結果とする音声データ識別手段とを有する。

【0027】[音声識別方法] また、本発明にかかる第1の音声識別方法は、話し声に含まれる1つ以上の単語を示す音声データから、前記単語それぞれを識別する音声識別方法であって、1つ以上の前記単語の音声データそれぞれに、これらの音声データの1つ以上それぞれを識別して得られる1つ以上の前記単語の組み合わせの候補(候補単語)の1組以上を対応付けるステップと、1つ以上の前記単語の音声データそれぞれに対応付けた前記候補単語それぞれに、前記候補単語それぞれの発音に対応しうる1つ以上の前記単語の組み合わせ(類似単語)の0組以上を対応付けるステップと、1つ以上の前記単語の音声データそれぞれに対応付けた前記候補単語、および、前記候補単語それぞれに対応付けた前記類似単語のいずれかを選択し、前記単語の音声データそれぞれの識別結果とするステップとを含む。

【0028】また、本発明にかかる第2の音声識別方法は、所定の言語以外を主に話す話者による前記所定の言語の音声に含まれる前記所定の言語の1つ以上の単語を示す音声データから、前記所定の言語の単語の1つ以上それぞれを識別する音声識別方法であって、前記所定の言語の単語の音声データの1つ以上それぞれに、これらの音声データの1つ以上それぞれを識別して得られる前記所定の言語の単語、および、前記話者が話した可能性がある前記所定の言語の単語の1つ以上またはこれらのいずれかを対応付けるステップと、1つ以上の前記単語の音声データそれぞれに対応付けた単語のいずれかを選択し、1つ以上の前記単語の音声データそれぞれの識別結果とするステップとを含む。

【0029】[音声矯正装置] また、本発明にかかる音声矯正装置は、単語を示す音声データを識別して得られる単語の候補(候補単語)を1個以上、対応付ける候補単語対応付け手段と、音声データに対応付けた前記候補単語それぞれに、前記候補単語それぞれの発音に対応しうる単語(類似単語)を0個以上、対応付ける類似単語対応付け手段と、前記音声データが示す単語と、この音声データに対応付けられた前記候補単語それぞれに対応付けられた前記類似単語とが一致する場合に、前記音声データが示す単語と同じ前記類似単語に対応し、前記音声データが示す単語の発音を矯正する発音矯正データを出力する発音矯正データ出力手段とを有する。

【0030】[発音矯正装置の作用] 本発明にかかる発音矯正装置において、候補単語対応付け手段および類似単語対応付け手段は、上述した本発明にかかる音声識別装置においてと同様に、音声データと、候補単語・類似単語とを対応付ける。

【0031】[発音矯正データ出力手段] 話者がネイティブに近い正しい発音をしている場合には、話者の意図した単語と音声データの識別結果とは候補単語に含まれることになる。一方、話者の発音が誤っていたり、不明瞭であったりすると、話者の意図した単語は候補単語に

含まれるが、音声データの識別結果は類似単語に含まれるということになる。従って、予め話者に発音すべき単語を示し、この単語を発音させた場合に、この単語が音声データの識別結果において、類似単語と一致すること、ユーザ（話者）の発音に何らかの誤り、あるいは、発音に不明瞭さがあることを意味する。

【0032】発音矯正データ出力手段は、話者に示した単語が類似単語と一致する場合に、話者に示した単語と一致する類似単語に対応付けられ、発音の誤り・不明瞭さを矯正する情報（例えば、ネイティブスピーカが正しい発音を行なう際の口および舌の動きを示す画像データ、および、ネイティブスピーカと比べて、話者の発音のどこが誤っているかを文章で示すテキストデータ）をモニタに表示し、話者に発音の矯正を促すとともに、話者の発音がネイティブスピーカの発音に近づくように学習を補助する。

【0033】[音声矯正方法] また、本発明にかかる音声矯正方法は、単語を示す音声データを識別して得られる単語の候補（候補単語）を1個以上、対応付け、音声データに対応付けた前記候補単語それぞれに、前記候補単語それぞれの発音に対応しうる単語（類似単語）を0個以上、対応付け、前記音声データが示す単語と、この音声データに対応付けられた前記候補単語それぞれに対応付けられた前記類似単語とが一致する場合に、前記音声データが示す単語と同じ前記類似単語に対応し、前記音声データが示す単語の発音を矯正する発音矯正データを出力する。

【0034】

【発明の実施の形態】 [第1実施形態] 以下、本発明の第1の実施形態を説明する。なお、以下、説明の明確化および便宜のために、特に断らない限り、本発明にかかる音声識別処理が、日本人が話す英語を識別するために調整されている場合を具体例として説明する。

【0035】[コンピュータ1] まず、図1を参照して、本発明にかかる音声識別処理を実現するコンピュータ1を説明する。図1は、本発明にかかる音声識別処理を実現するコンピュータ1の構成を例示する図である。図1に例示するように、コンピュータ1は、例えば、音声入出力機能を有するパーソナルコンピュータであって、CPU、メモリおよびこれらの周辺装置等を含むコンピュータ本体10、出力装置100、光磁気(mo)ディスク装置、ハードディスク装置あるいはフロッピーディスク装置等の記憶装置110、および、入力装置120から構成される。出力装置100は、CRディスプレイ装置等のモニタ102、音声出力用のスピーカ104、プリンタ106等を含む。入力装置120は、マイク122、音声入力用ボード124、キーボード126およびマウス128等を含む。

【0036】[モニタ102] 出力装置100において、モニタ102は、コンピュータ1のユーザに対して

操作用のGUI画像、および、コンピュータ本体10が音声データを識別して得られたテキストデータ等を表示する。

【0037】[スピーカ104] スピーカ104は、コンピュータ本体10が音声データを識別して得られたテキストデータを音声として出力するため等に用いられる。

【0038】[プリンタ106] プリンタ106は、コンピュータ本体10が音声データを識別して得られたテキストデータのハードコピーを出力するため等に用いられる。

【0039】[記憶装置110] 記憶装置110は、コンピュータ本体10による制御に従って動作し、コンピュータ本体10が音声データを識別して得られたテキストデータを記憶する。また、記憶装置110は、音声識別に必要なデータ（以下、このようなデータを「単語データ」と総称する）およびプログラム等を記憶し、記憶した単語データおよびプログラム等をコンピュータ本体10に対して出力する。なお、記憶装置110に記憶される単語データは、例えば、音声識別プログラム16により作成され、あるいは、フロッピーディスク等の記録媒体により供給され、ラベル列テーブル、インデックステーブル、単語レコード、類似単語レコードおよび誤り情報コードテーブル（ラベル列データを除くこれらの詳細は、図4～図7を参照して後述する）を含む。

【0040】[マイク122] マイク122は、ユーザが発音した話し声の音声を集音し、アナログ形式の音声信号に変換して音声入力用ボード124に対して出力する。

【0041】[音声入力用ボード124] 音声入力用ボード124は、コンピュータ本体10による制御に従って動作し、マイク122から入力された音声信号をサンプリングし、音声信号の波形に対応するデジタル形式の音声データに変換してコンピュータ本体10に対して出力する。

【0042】[キーボード126、マウス128] キーボード126およびマウス128は、例えば、モニタ102に表示されたGUIに対するユーザの操作を受け入れて、操作入力としてコンピュータ本体10に対して出力する。

【0043】[ソフトウェア14] 以下、図2を参照して、本発明にかかる音声識別処理を実現するソフトウェアの構成を説明する。図2は、本発明にかかる音声識別処理を実現するソフトウェア14の構成を示す図である。なお、図2においては、本発明にかかる音声識別処理の実現に関係がないソフトウェアの構成部分を省略してある。

【0044】図2に示すように、ソフトウェア14は、ハードウェア(H/W) サポート部142、オペレーティングシステム(OS) 148およびアプリケーション部から構成される。ハードウェアサポート部142は、音声デバイスドライバ144および記憶デバイスドライバ146を含む。オペレーティングシステム148は、例

13

えば、OS/2 (IBM社商品名) あるいはWindows (マイクロソフト社商品名) といった汎用OSであって、音声インターフェース (IF) 部150および記憶装置インターフェース部152を含む。また、ソフトウェア14は、アプリケーション部として音声識別プログラム16を含む。ソフトウェア14のこれらの構成部分は、記憶装置110に記憶され、必要に応じてコンピュータ本体10のメモリにロードされて実行される。

【0045】[音声デバイスドライバ144] ハードウェアサポート部142において、音声デバイスドライバ144は、音声入力用ボード124を制御して、マイク122から入力される音声信号を音声データに変換させる。また、音声デバイスドライバ144は、音声入力用ボード124から入力された音声データを音声インターフェース部150に対して出力するインターフェース機能を実現する。また、音声デバイスドライバ144は、オペレーティングシステム148の音声インターフェース部150の制御に従って、音声入力用ボード124のサンプリング周期の変更等の設定変更、および、サンプリングの開始および終了等の動作制御を行なう。

【0046】[記憶デバイスドライバ146] 記憶デバイスドライバ146は、オペレーティングシステム148の記憶装置インターフェース部152からの要求 (制御) に応じて、記憶装置110の動作を制御し、単語データおよび音声識別の結果として得られたテキストデータを記憶させ、あるいは、記憶装置110が記憶しているこれらのデータを読み出させる。また、記憶デバイスドライバ146は、記憶装置110から入力された単語データおよびテキストデータを、記憶装置インターフェース部152に対して出力し、あるいは、記憶装置インターフェース部152から入力されたこれらのデータを記憶デバイスドライバ146に対して出力するインターフェース機能を実現する。

【0047】[オペレーティングシステム148] オペレーティングシステム148は、音声インターフェース部150および記憶装置インターフェース部152が実現する機能の他、コンピュータ本体10におけるプログラムの実行制御を行なう。また、オペレーティングシステム148は、音声識別プログラム16が出力するテキストデータおよびGUI画像をモニタ102に表示する処理、テキストデータを音声信号に変換してスピーカ104を介して出力する処理、プリンタ106に対するハードコピーを行なうために必要な処理、および、キーボード126およびマウス128に対するユーザの操作を受け入れる等を行なう。

【0048】[音声インターフェース部150] オペレーティングシステム148において、音声インターフェース部150は、音声識別プログラム16からの要求 (制御) に応じて、音声デバイスドライバ144を制御する。また、音声インターフェース部150は、音声デ

14

バイスドライバ144から入力された音声データを音声識別プログラム16に対して出力するインターフェース機能を実現する。

【0049】[記憶装置インターフェース部152] 記憶装置インターフェース部152は、記憶装置110の記憶領域を管理する。また、記憶装置インターフェース部152は、音声識別プログラム16からの要求 (制御) に応じて記憶デバイスドライバ146を制御し、音声識別プログラム16から要求された単語データおよびテキストデータを記憶装置110から読み出させ、読み出された単語データおよびテキストデータを音声識別プログラム16に対して出力する。また、記憶装置インターフェース部152は、音声識別プログラム16から入力される単語データおよびテキストデータを、記憶デバイスドライバ146を介して記憶装置110の空き記憶領域に記憶させる。

【0050】[音声識別プログラム16] 以下、図3を参照して音声識別プログラム16を説明する。図3は、図2に示した音声識別プログラム16の構成を示す図である。

【0051】図3に示すように、音声識別プログラム16は、単語データベース部160、制御部162および音声識別部18から構成される。音声識別部18は、ベクトルデータ生成部180、ラベル作成部182、候補単語作成部184、類似単語追加部186および絞り込み部188を含む。音声識別プログラム16は、これらの構成部分により、操作のGUI画像を表示し、表示したGUI画像に対するユーザの操作に従って、音声インターフェース部150から入力される音声データを、記憶装置インターフェース部152から入力される単語データを用いて識別し、識別の結果として得られた単語列をテキストデータとしてオペレーティングシステム148を介して出力する。

【0052】[制御部162] 制御部162は、操作のGUI画像をモニタ102に表示し、表示したGUI画像に対して、ユーザが入力装置120のキーボード126およびマウス128を用いて行なう操作をオペレーティングシステム148を介して受け入れる。また、制御部162は、受け入れた操作入力に応じて、オペレーティングシステム148の音声インターフェース部150および記憶装置インターフェース部152を制御する。

【0053】また、制御部162は、受け入れた操作入力に応じて単語データベース部160を制御して、候補単語作成部184が音声データと候補単語とを対応付けるために用いるラベル列テーブル、類似単語追加部186が類似単語を候補単語と対応付けるために用いるインデックステーブル、単語レコード、類似単語レコードおよび誤り情報コードテーブル (図4～図7を参照して後述する) を含む単語データを作成あるいは更新させ、記

憶装置インターフェース部152等を介して記憶装置110に記憶させる。

【0054】また、制御部162は、GUI画像内に、音声データの各部分と対応付けた候補単語および類似単語を表示し、表示したこれらの単語に対する操作入力に応じて、候補単語および類似単語のいずれかを絞り込み部188に選択させ、最終的な識別結果とさせる。なお、制御部162による候補単語および類似単語の表示方法の例として、候補単語作成部184および類似単語追加部186が対応付けた候補単語をモニタ102に反転表示し、ユーザのキーボード126に対する操作に応じて候補単語および類似単語を変更して順次、表示する方法、あるいは、ユーザがモニタ102に表示された候補単語の誤り部分を見つけて、マウス128でクリックした場合に、クリックされた部分の候補単語と対応付けたウィンドウ内に、候補単語および類似単語の一覧を表示する等を挙げることができる。

【0055】[単語データベース部160] 単語データベース部160は、上述のように制御部162の制御に従って単語データを作成または更新し、記憶装置110に記憶させ、管理する。また、単語データベース部160は、候補単語作成部184に対して単語データ（ラベル列テーブル）を出力する。

【0056】また、単語データベース部160は、類似単語追加部186の要求に応じて単語データ（インデックステーブル、単語レコード、類似単語レコードおよび誤り情報コードテーブル；図4～図7）を検索し、検索の結果として得られ、類似単語追加部186に入力された第1の候補単語に対応する単語レコード、類似単語レコードおよび誤り情報を類似単語追加部186に対して出力する。

【0057】[単語データ] ここで、以下の説明の理解を容易にするために、図4～図7を参照して、ラベル列テーブルを除く単語データ（インデックステーブル、単語レコード、類似単語レコードおよび誤り情報テーブル）を説明する。

【0058】[インデックステーブル] 図4は、単語データのインデックステーブルに含まれるデータを例示する図である。インデックステーブルは、単語データベース部160が、単語の先頭文字（A～Z）により分類された単語レコードを検索するために用いられる。インデックステーブルは、図4に示すように、先頭文字A～Zそれぞれの記録領域の先頭を示すポインタと、先頭文字をそれぞれA～Zとする単語レコードの数とが対応付けられて構成される。

【0059】[単語レコード] 図5は、単語データの単語レコードに含まれるデータを例示する図である。図5に示すように、単語レコードは、類似単語追加部186において、候補単語作成部184が作成した第1の候補単語データ〔候補単語データ（1）〕と突き合わされ、

比較される見出し単語（TarWord）、次の見出し単語へのポインタ（NextP）、単語レコードに含まれる類似単語の数（#Can）および類似単語レコード（CanWord）が対応付けられて構成される。

【0060】なお、図5においては、類似単語レコードが単語レコードに直接、対応付けた場合が示されているが、例えば、単語レコードに類似単語レコードのポインタを対応付け、類似単語レコードを単語レコードと別ファイルとする等、実現方法は問わない。また、単語レコードそれぞれの見出し単語（TarWord）を複数にして、類似単語追加部186が、連続した複数の単語に対応するラベルと、複数の単語を含む見出し単語とを対応付けられるようにしてもよい。

【0061】[類似単語レコード] 図6は、単語データの類似単語レコードに含まれるデータを例示する図である。図6に示すように、類似単語レコードは、入力単語数（#m；#mは1以上の整数）、入力候補単語（aWord、aWord-1、aWord-2、...、aWord-m-1）、出力単語数（#n；#nは0以上の整数）、類似単語（COWord、COWord-1、...、COWord-n）および誤りコード（ECode）が対応付けられて構成される。

【0062】これらの内、入力単語数（#m）は、類似単語追加部186が、候補単語作成部184から入力された第1の候補単語を類似単語に対応付ける際に、前後いくつかの候補単語を参照するかを示す（以下、第1の候補単語の後方のm-1文字を参照する場合を例として説明する）。

【0063】入力候補単語（aWord、aWord-1、aWord-2、...、aWord-m-1）は、候補単語作成部184から類似単語追加部186に連続して入力されたm個の第1の候補単語（TarWord、TarWord-1、...、TarWord-m-1）と突き合わされ、比較される単語列を示す。つまり、類似単語追加部186においては、第p番目の第1の候補単語（TarWord）は、入力されても直ちに類似単語と対応付けられず、さらに類似単語追加部186にm-1個の第1の候補単語が入力された後に、第p番目～第p+m-1番目のm個の連続した第1の候補単語（TarWord、TarWord-1、...、TarWord-m-1）それぞれと、類似単語レコードのm個の入力候補単語（aWord、aWord-1、aWord-2、...、aWord-m-1）それぞれとが比較され、これらが一致した場合にのみ、第p番目の第1の候補単語（aWord = TarWord）と、類似単語レコード内で入力候補単語に続くn個の類似単語（COWord、COWord1、...、COWord-n）とが対応付けられる。なお、類似単語が存在しない場合には、出力単語数（#n）の値は0とされ、類似単語は類似レコード内に対応付けられない。

【0064】図6に示した類似単語レコードにおける第1の候補単語と類似単語との間のマッピング方法について、さらに説明する。第1の候補単語と類似単語との間の対応付け（マッピング）の方法としては、例えば、以

下に示す4つの方法が考えられる。

【0065】[第1の方法] 1つの単語が別の1つの単語に誤って識別される場合に対応するために、1つの単語のみを含む第1の候補単語に、1つの単語のみを含む類似単語を対応付ける。第1の方法の例としては、第1の候補単語“read”の“r”が、正しく発音されなかった場合に備えて、第1の候補単語“read”と類似単語“lead”とを対応付けることが挙げられる。さらに、第1の方法をとる場合の対応付けの例としては、“sink”と“think”、“fell”と“fill”、“seat”と“sit”、“better”と“bitte”、“nut”と“not”、“fund”と“found”、“boat”と“bought”および“coal”と“call”との対応付け等を挙げることができる。

【0066】[第2の方法] 1つの単語が別の複数の単語に誤って識別される場合に対応するために、1つの単語のみを含む第1の候補単語に、複数の単語を含む類似単語を対応付ける。第2の方法の例としては、第1の候補単語“jumped”の“ed”が、正しく“t”と発音されなかった場合に備えて、第1の候補単語“jumped”と類似単語“jump”、“and”を対応付けることが挙げられる。さらに、第2の方法をとる場合の対応付けの例としては、“check in”と“chicken”との対応付け等を挙げることができる。

【0067】[第3の方法] 複数の単語が別の1つの単語に誤って識別される場合に対応するために、複数の単語を含む第1の候補単語に、1つの単語のみを含む類似単語を対応付ける。第3の方法の例としては、第1の候補単語“have”、“to”がつながって発音された場合に備えて、第1の候補単語“have”、“to”と類似単語“hat”を対応付けることが挙げられる。さらに、第3の方法をとる場合の対応付けの例としては、“I will”と“aisle”との対応付け等を挙げることができる。

【0068】[第4の方法] 複数の単語が別の複数の単語に誤って識別される場合に対応するために、複数の単語を含む第1の候補単語に、複数の単語を含む類似単語を対応付ける。第1～第3の方法は、第4の方法に限定を加えた方法と考えることができるので、図6に示した類似単語レコードは、第4の方法に基づいて作成され、複数の単語を含む第1の候補単語に複数の単語を含む類似単語が対応付けてある。

【0069】類似単語は、コンピュータ1を利用する話者(ユーザ)の発音が不正確な場合に、候補単語の代わりに選択される。従って、絞り込み部188においては、候補単語ではなく、類似単語が最終的に選ばれた場合には、話者は、英語の発音上、選ばれた類似単語に対応する誤りをしていることになる。誤りコード(ECode)は、このような観点から類似単語レコードに付加され、最終的に選択された類似単語に対応する発音上の誤りを符号の形式で示す。

【0070】[誤り情報コードテーブル]図7は、単語データの誤り情報コードテーブルを例示する図である。図

7に示すように、誤り情報コードテーブルは、誤りコード(ECode: 0, 1, 2, ...)と、誤りの内容を示す情報(例えば、「rをlと発音した」、「lをrと発音した」、「thをsと発音した」等の誤り情報)とが対応付けられて構成される。

【0071】[ベクトルデータ生成部160]音声識別部18において、ベクトルデータ生成部160(図3)は、音声インターフェース部150から入力される音声データを処理して、音声の複数の特徴(音の高さ、強さ、イントネーション等)それぞれについて量子化を行ない、これらの特徴それぞれを示す数値を含むベクトルデータを生成して制御部162に対して出力する。例えば、音声データのサンプリング周波数が11kHzである場合、ベクトルデータ生成部160は、音声データを1/100秒単位で処理し、音声データの複数の種類の特徴それぞれを量子化し、複数の要素からなるベクトルデータを生成する。

【0072】[ラベル作成部182]ラベル作成部182は、ベクトルデータ生成部160から入力されたベクトルデータを、発音記号に類似するラベルと呼ばれるデータに変換し、1センテンス分ずつ候補単語作成部184に対して出力する。ラベル作成部182は、この変換処理を、例えば、色々な人(大人、子供、男、女等)の実際の話し声のサンプルから生成したラベルと、連続した複数のベクトルデータのパターンとを対応付けたラベルテーブルを用い、連続した複数のベクトルデータに対応するラベルを選択することにより行なう。ただし、「センテンス」という用語は、実際の文章内の実際のセンテンスに必ずしも対応せず、単に音声識別の処理単位を示す。

【0073】[候補単語作成部184]候補単語作成部184は、音声データにおいて、1つ以上の英単語に対応する連続した1つ以上のラベル列を、英単語との対応を示すラベル列テーブルを用いて、ラベル列それぞれが示す1つ以上の英単語の組み合わせの1つ以上と対応付け、対応付けた英単語の組み合わせを第1の候補単語として類似単語追加部186に対して出力する。(以下、説明の簡略化のために、候補単語作成部184が、1つの英単語に対応するラベル列それぞれを、1つの英単語のみを含む第1の候補単語に対応付ける場合を例として説明する。)

【0074】ここで、例えば、候補単語作成部184は、ラベルが示す音をアルファベットに変換し、変換して得られたアルファベット列を英単語に変換するのではなく、ラベル列を直接、英単語(第1の候補単語)に変換する。つまり、候補単語作成部184は、例えば、第1の候補単語として“read”を作成する際に、ラベル列を“r”、“e”、“a”、“d”という4個のアルファベットに置換してから“read”という単語を第1の候補単語として対応付けるのではなく、ラベル列に直接に“read”という単語

語を対応付ける。

【0075】なお、候補単語作成部184は、作成した第1の候補単語を、図8に示す入力レコード(InWord)、および、図9に示す入力レコードマトリクス(InMatrix)の形式で、1センテンス分ずつ類似単語追加部186に対して出力する。図8は、候補単語作成部184が類似単語追加部186に出力する入力レコード(InWord)のデータ構造を示す図である。図9は、候補単語作成部184が類似単語追加部186に出力する入力レコードマトリクス(InMatrix)のデータ構造を示す図である。

【0076】図8に示すように、候補単語作成部184は、ラベル列と対応付けた単語およびその単語長を示すデータ(InWord)それぞれに、その単語が1つのセンテンスにおいて第i番目であること、および、1つのセンテンスの第i番目の第j番目の第1の候補単語であることを示すデータを付加し、入力レコード(InWord)を作成して類似単語追加部186に対して出力する。ただし、i, jは整数であって、iは最大単語数(Maxi)以下、jは最大候補数(Maxj)以下である。

【0077】さらに、候補単語作成部184は、図8に示すように、ラベル作成部182から入力されたラベル列と、選択した英単語に対応するラベル列テーブル内のラベル列との一致の程度、言いかえると、ラベル列が第1の候補単語を示している確率を示す確率データを作成し、入力レコードの単語およびその単語長を示すデータに付加して類似単語追加部186に対して出力する。

【0078】また、1センテンス分の入力レコード(InWord)の作成が終了すると、候補単語作成部184は、図9に示すように、そのセンテンス中に含まれる最大単語数(Maxi)、同一のラベル列(読み)に対して、最大いくつかの第1の候補単語が対応付けられたかを示す最大候補数(Maxj)、および、第i番目の単語に第j番目の単語が存在するかを示すフラグFlg(i, j)を示す入力レコードマトリクスを作成し、1センテンス分の入力レコードとともに類似単語追加部186に対して出力する。なお、候補単語作成部184が第i番目の単語に対応するラベル列の第1の候補単語を選ばなかった場合には、フラグFlg(i, 1)は、第i番目の第1番目の単語が存在しない旨を示す値(例えば0)とされる。

【0079】〔類似単語追加部186〕類似単語追加部186は、候補単語作成部184から入力された入力レコードそれぞれに対して、ラベル列と対応付けることはできないが、日本人による英語の発音の癖等を考慮して、第1の候補単語に類似する可能性がある英単語を第1の候補単語に付加し、第2の候補単語を生成して絞り込み部188に対して出力する。

【0080】さらに類似単語追加部186の動作を詳細に説明する。類似単語追加部186は、まず、候補単語作成部184から入力された1センテンス分の入力レコード(InWord)に含まれる第p番目の第1の候補単語を順

次、単語データベース部160に対して出力し、単語レコードの取得を要求する。単語データベース部160は、インデックステーブル(図4)を用いて単語レコード(図5)の検索を行ない、入力レコード(InWord)に含まれる単語と見出し単語(TarWord)が一致する単語レコードを取得し、類似単語追加部186に対して出力する。

【0081】類似単語追加部186は、第p番目の第1の候補単語の単語インデックスを得ると、第p番目から第p+m-1番目の単語(InWord-p, j, InWord-p+1, j, ..., InWord-p+m-1, j)と、単語データベース部160から入力される単語レコードに付加された類似単語レコード(図6)それぞれのm個の入力単語(aWord, aWord-1, aWord-2, ..., aWord-m-1)とを比較し、これらが一致した場合には、n個の類似単語(COWord-1, COWord-2, ..., COWord-n)を第p番目の第1の候補単語に付加して第2の候補単語を作成する。

【0082】なお、類似単語追加部186が、第1の候補単語に類似単語を付加して第2の候補単語を作成するのではなく、第1の候補単語を類似単語で置き換えて第2の候補単語を作成するように処理を変更することも可能である。この場合には、類似単語に第1の候補単語が含まれているか否かを問わない。

【0083】さらに、類似単語追加部186は、第1の候補単語に付加した類似単語レコードに対応する誤りコード(ECode)を単語データベース部160に対して出力し、誤りコードが示す誤り情報の取得を要求する。単語データベース部160は、この要求に応じて誤り情報コードテーブル(図7)を検索し、誤り情報を取得して類似単語追加部186に対して出力する。

【0084】図10は、類似単語追加部186が絞り込み部188に出力する出力レコード(OutWord)のデータ構造を示す図である。図11は、類似単語追加部186が絞り込み部188に出力する出力レコードマトリクス(OutMatrix)のデータ構造を示す図である。

【0085】なお、類似単語追加部186は、第2の候補単語、単語長を示すデータ、確率データおよび誤り情報(または誤り情報コード)と、その単語が1つのセンテンスにおいて第i'番目であること、および、1つのセンテンスの第i'番目の第j'番目の第1の候補単語であることを示すデータを付加し、図10に示すように、入力レコード(図8)と同様な形式をとる出力レコード(OutWord)の形式で絞り込み部188に対して出力する。

【0086】また、1センテンス分の出力レコード(OutWord)の作成が終了すると、類似単語追加部186は、図11に示すように、入力レコードマトリクス(図9)と同様に、そのセンテンス中に含まれる最大単語数(Maxi')、同一のラベル列(読み)に対して、最大いくつかの第2の候補単語が対応付けられたかを示す最大候補数(M

axj'), および、第 i' 番目の単語に第 j' 番目の単語が存在するかを示すフラグ Fig(i' j') を示す出力レコードマトリクスを作成し、1 センテンス分の出力レコードとともに絞り込み部 188 に対して出力する。

【0087】 [絞り込み部 188] 絞り込み部 188 は、出力レコードとして類似単語追加部 186 から入力された第 2 の候補単語をモニタ 102 に表示し、例えば、ユーザの操作に応じて、あるいは、それまでに識別した単語列の構文解析結果に基づいて、第 2 の候補単語のいずれかを、最終的な識別結果として選択し、選択した単語を並べたテキストデータを作成し、モニタ 102、スピーカ 104 あるいはプリンタ 106 に対して出力する。

【0088】 絞り込み部 188 によるテキストデータの作成についてさらに説明する。例えば、1 つのセンテンスの第 1 番目の単語の第 2 の候補単語が n1 個 (OutWord-1.1, OutWord-1.2, ..., OutWord-1.n1)、第 2 番目の単語の第 2 の候補単語が n2 個といったようにある場合には、絞り込み部 188 は、第 1 番目の単語の第 2 の候補単語のいずれか、および、第 2 番目の単語の第 2 の候補単語のいずれかといったように、各単語の第 2 の候補文字のいずれかを、単語の順番通りにモニタ 102 に表示する。

【0089】 ユーザが、例えば、マウス 128 を用いて GUI 画像のウィンドウ内に表示された第 2 の候補単語をクリックすると、制御部 162 は、クリックされた第 2 の候補単語を反転表示し、その後、ユーザが同じ部分をクリックするたびに、表示する第 2 の候補単語を変更する。ユーザが最終的に第 2 の候補単語を選択し、その旨の操作をマウス 128 あるいはキーボード 126 に対して行なったり、あるいは、次の単語の選択操作に移行すると、絞り込み部 188 は、最後に表示した第 2 の候補単語を最終的に選択された識別結果とする。ユーザは必要に応じてこの作業を繰り返し、絞り込み部 188 は、この作業に応じて単語の選択を行ない、テキストデータを作成する。

【0090】 なお、絞り込み部 188 の処理を、第 2 の候補単語を表示する際に、単語の前後関係を考慮して、最終的な識別結果として選択される可能性が高い第 2 の候補単語のみをモニタ 102 に表示したり、あるいは、最終的な識別結果として選択される可能性が高い順番でモニタ 102 に表示させるようにしたりすると便利である。

【0091】 便宜的に日本語を識別する場合を具体例として挙げて説明する。候補単語作成部 184 は、「あかい はな」という日本語の音声データから得られたラベルを、「たかい・はな (高い・花, 高い・鼻)」と識別し、さらに、類似単語追加部 186 がこれらに類似単語「赤い・鼻」を付加して、これら 3 種類を第 2 の候補単語として絞り込み部 188 に出力する可能性がある。こ

のような場合、絞り込み部 188 において、第 2 の候補単語の前半の識別結果が「赤い」とされた場合に、後半の識別結果は、「花」、「鼻」および「鼻」の順で確からしいと考えられる。このような場合、絞り込み部 188 の処理を、「赤い」という単語に続けて「花」、「鼻」のみを表示するようにしたり、「花」、「鼻」、「鼻」の順番で表示するようにすると、ユーザの選択操作を最小限にすることができる。

【0092】 また、絞り込み部 188 の処理を、類似単語追加部 186 から入力される出力レコードに付加された確率データを利用して、例えば、ユーザが設定する閾値以上の範囲の値をとる確率データに対応する第 2 の候補単語のみをモニタ 102 に表示するようにすると、最終的な識別結果として選択される可能性が高い第 2 の候補単語のみがモニタ 102 に表示されることとなり、ユーザの選択操作をより少なくすることができる。

【0093】 [コンピュータ 1 の動作] 以下、図 12 を参照して、コンピュータ 1 における音声識別処理を説明する。図 12 は、コンピュータ 1 における本発明にかかる音声識別処理を示すフローチャート図である。なお、図 12 においては、説明の簡略化のために、本発明にかかる音声識別処理の内、基本的な処理のみを示してあり、上述した確率データあるいは誤り情報を用いた処理は適宜、省略されている。

【0094】 図 12 に示すように、ステップ 100 (S100) において、音声識別プログラム 16 のベクトルデータ生成部 160 およびラベル作成部 182 (図 3) は、日本人ユーザが発音した 1 センテンス (単位) 分の音声データを処理し、ラベル列に変換する。ベクトルデータ生成部 160 およびラベル作成部 182 が生成した 1 センテンス分のラベル列は、候補単語作成部 184 に入力される。

【0095】 ステップ 102 (S102) において、候補単語作成部 184 は、ラベル作成部 182 から入力された 1 センテンス分のラベル列を第 1 の候補単語に対応付け、図 8 に示した入力レコード (InWordij) の形式で類似単語追加部 186 に対して出力し、さらに、図 9 に示した入力レコードマトリクス (InMatrix) を作成して類似単語追加部 186 に対して出力する。

【0096】 ステップ 104 (S104) において、類似単語追加部 186 は、単語データベース部 160 に対して処理の対象となっている入力レコードに含まれる第 1 の候補単語の単語レコード (図 5) の検索を要求する。単語データベース部 160 は、類似単語追加部 186 の要求に応じてインデックステーブル (図 4) を用いて検索を行ない、検索の結果として第 1 の候補単語 (入力レコード) に対応する単語レコードが得られた場合には、得られた単語レコードを類似単語追加部 186 に対して出力して S106 の処理に進み、得られなかった場合には S110 の処理に進む。

【0097】ステップ106 (S106) において、類似単語追加部186は、単語データベース部160から入力された単語レコードの類似単語レコード (図6) を処理し、第1の候補単語 (入力レコード) に対応する類似単語を取得する。

【0098】ステップ108 (S108) において、取得した類似単語を第1の候補単語に付加して第2の候補単語を作成する。

【0099】ステップ110 (S110) において、類似単語追加部186は、1センテンスに含まれる入力レコードの全ての処理を終了したか否かを判断し、終了した場合にはS110の処理に進み、終了しない場合には、処理対象を次の入力レコードに変更してS104の処理に戻る。

【0100】ステップ112 (S112) において、類似単語追加部186は、S108において作成された第2の候補単語を、図10に示した出力レコードの形式で絞り込み部188に対して出力する。さらに、類似単語追加部186は、第2の候補単語に対応する出力レコードマトリクス (図11) を作成し、絞り込み部188に

対して出力する。絞り込み部188は、入力された第2の候補単語をモニタ102上のGUI画像のウィンドウ内に表示し、ユーザの操作に応じて最終的な識別結果を、英文のテキストデータとして出力する。

【0101】〔変形例〕なお、絞り込み部188の処理を、ユーザが、類似単語追加部186において類似単語として第2の候補単語に付加された単語を最終的な識別結果として選択した場合に、選択された単語に付加された誤り情報をモニタ102に表示するようにすると、ユーザは、自分の英語の発音の欠点を知ることができ、コンピュータ1を音声識別装置として用いるほか、英語の発音の学習装置として用いることができる。誤り情報の表示方法としては、正しい発音をするための口の形を表示する、あるいは、正しい発音を音声合成してユーザに聞かせる等の方法が考えられる。

【0102】また、上述した第1の実施形態においては、日本人が発音した英語から英文テキストを識別する処理を例示したが、本発明にかかる音声識別処理は、このような場合に限定されず、例えば、アメリカ式英語を話すアメリカ人が、英国式英語用に調整された音声識別装置を用いてテキストを作成する等、同一の言語であっても発音が異なる場合、あるいは、個人の発音に癖があったり、不明りょうであったりする場合の識別率の低下に対処するために、広く応用することができる。

【0103】以上述べたように、本発明にかかる音声識別処理においては、候補単語作成部184が識別した第1の候補単語に、類似単語追加部186において類似単語を付加して第2の候補単語とするので、音声識別率が向上する。便宜的にコンピュータ1により日本語を識別する場合を具体例として説明する。例えば、ユーザが日

本語で「あかいはな (赤い花)」と発音したつもりであっても、「あ」の発音が何らかの原因で不明りょうになって、ラベル作成部182が、「あかいはな」とも「たかいはな」ともつかないラベル列を生成することがある。

【0104】候補単語作成部184が、このラベル列に対応する第1の候補単語として「高い・鼻」を選択し、類似単語追加部186に出力した場合であっても、単語レコードが適切に作成されていれば、類似単語追加部186が第1の単語として「赤い・鼻」、 「赤い・花」を第1の候補単語に付加して第2の候補単語とすることができる。従って、候補単語作成部184が作成した第1の候補単語には含まれていなかった本来の「赤い・花」もモニタ102に表示されることになり、ユーザは第2の候補単語の中から正しい識別結果を選択することができる。

【0105】また、通常、音声識別方法は、純粋な音声処理に文法的解析処理 (文章中の特定の位置には名詞がおかれやすい等) や、言語モデル処理 (並んだ単語の確からしさ) 等が組み合わされて実現されるので、これらの処理を行なう前に、本発明にかかる音声識別方法で類似単語を追加しておく、音声識別率が大幅に向上するなど、効果が非常に大きい。

【0106】〔第2実施形態〕図3に示した音声識別プログラム16において、音声識別の結果として候補単語が選択されるということは、ユーザ (話者) の発音がネイティブスピーカに比較的近く、本発明にかからない一般的な音声識別装置を用いても高い識別率で識別可能な程度に正確であることを意味し、反対に、類似単語が選択されるということは、話者の発音に誤り、あるいは、不明瞭さがあることを意味する。このことは、音声識別プログラム16において、ユーザが意図した単語が類似単語に含まれる場合にも当てはまる。

【0107】従って、ユーザが意図した単語が、類似単語として識別された場合に、話者に対して発音がどのように誤っているか、正しく発音するにはどのようにしたらよいか等の情報 (発音矯正情報) を提示することにより、話者の発音の学習を補助でき、話者の発音を矯正できることが分かる。以下、第2の実施形態として示す発音矯正方法は、この点に着目し、第1の実施形態として示した音声識別処理を変形し、話者の発音が不正確であると判断できる場合に、話者に発音矯正情報を示す画像表示するようにしたものである。

【0108】〔コンピュータ2〕図13は、本発明にかかる音声識別処理および発音矯正方法を実現するコンピュータ2の構成を示す図である。なお、特に断らない限り、以下の図面に示す構成部分は、これまでの図面に示した同一符号の構成部分と同じである。図13に示すように、コンピュータ2は、コンピュータ1 (図1) の入力装置120を入力装置130で置換した構成を採り、

入力装置130は、入力装置120に画像入力用ボード132を追加した構成を採る。画像入力用ボード132は、例えば、発音矯正画像(図19)に用いられる画像データをビデオカメラとり込むために用いられる。

【0109】[ソフトウェア20]図14は、本発明にかかる音声識別処理および発音矯正方法を実現するソフトウェア20を示す図である。図14に示すように、ソフトウェア20は、ソフトウェア14(図2)における音声識別プログラム16を音声識別・矯正プログラム22で置換した構成を採る。ソフトウェア14において各構成部分の間で入出力されるデータの他に、ソフトウェア20においては、画像データがさらに入出力され、ソフトウェア14における識別結果(テキストデータ)の代わりに、ユーザ(話者)の発音を矯正する発音矯正情報を示す画像(発音矯正画像)がモニタ102等に出力されるようになっている。

【0110】[音声識別・矯正プログラム22]図15は、図14に示した音声識別・矯正プログラム22の構成を示す図である。図15に示すように、音声識別・矯正プログラム22は、音声識別プログラム16(図3)および発音矯正プログラム24から構成される(絞り込み部188は省略)。

【0111】[類似単語追加部186の変更点]音声識別・矯正プログラム22においては、音声識別プログラム16においてと異なり、類似単語追加部186は、発音矯正プログラム24の比較部240および発音矯正情報表示部242に対して類似単語レコード(図18)を出力する。

【0112】[制御部162の変更点]図16は、図15に示した制御部162が表示する発音指示画像を例示する図である。図17(A)、(B)はそれぞれ、図15に示した単語データベース部160が生成する矯正情報インデックスを例示する図であって、(A)はrの発音を矯正するための発音矯正画像を示し、(B)はlの発音を矯正するための発音矯正画像を示す。図18は、第2の実施形態において単語データベース部160が生成する類似単語レコードを示す図である。

【0113】制御部162は、図16に例示するように、ユーザに発音すべき単語(図16においては“read”)を示し、発音を促す発音指示画像(図16においては“read”と発音してみてください!)というテキストデータを含む画像)をさらに生成してモニタ102に表示し、ユーザに発音を指示した単語(正解単語)を比較部240に対して出力する。また、制御部162は、発音矯正画像表示部242から矯正情報コード(CCode;図18)が入力された場合に、この矯正情報コードを単語データベース部160に対して出力し、この矯正情報コードが示す矯正情報インデックス(図17(A)、

(B))を単語データベース部160から得る。さらに、制御部162は、この矯正情報インデックスに含ま

れるn個のエントリ(nは整数;図17(A)、(B)においてはn=8)が示す画像データおよびテキストデータを記憶装置110から読み出して、発音矯正画像表示部242に対して出力する。

【0114】図19は、図17(A)に例示した矯正情報インデックスが示す第1の発音矯正画像を例示する図である。なお、図19においては、図示の簡略化のために、図17(A)に例示したエントリ5~7に対応するテキストデータは省略されている。また、制御部162は、図17(A)に例示したような矯正情報インデックスに含まれるエントリと、図19に例示するようなエントリそれぞれが示す画像データ(Image)およびテキストデータ(Text)とを対応付けて、記憶装置110に記憶する。

【0115】図17(A)に例示した矯正情報インデックスは、ユーザのrの発音を矯正するために用いられ、発音の矯正すべき点を示すテキストデータ、rを発音する際の口の形を示す画像データ、lを発音する際の口の形を示す画像データ、rを発音するためのアドバイスを示すテキストデータ、rを含む単語の例、lを含む単語の例、rとlとが現れる単語の例を示すテキストデータ、および、rとlとが現れる文章の例を示すテキストデータをそれぞれ示すエントリ(エントリ1~4, 8)を含む。図17(A)に例示した矯正情報インデックスからは、図19に例示するような発音矯正画像が生成され、モニタ102に表示される。

【0116】また、図17(B)に例示した矯正情報インデックスは、ユーザのlの発音を矯正するために用いられ、発音の矯正すべき点を示すテキストデータ、lを発音する際の口の形を示す画像データ、rを発音する際の口の形を示す画像データ、lを発音するためのアドバイスを示すテキストデータ、lを含む単語の例、rを含む単語の例、rとlとが現れる単語の例を示すテキストデータ、および、rとlとが現れる文章の例を示すテキストデータをそれぞれ示すエントリを含む。

【0117】[単語データベース部160の変更点]単語データベース部160は、ソフトウェア14においてと異なり、誤り情報コードテーブル(図7)の代わりに、図17(A)、(B)に例示したような矯正情報インデックスを作成し、単語データとして記憶装置110にさらに記憶する。

【0118】また、単語データベース部160は、図18に示すように、矯正情報インデックスのいずれかを示す矯正情報コード(CCode; correction code)を、誤りコード(ECode;図6)の代わりに類似単語レコードに付加し、記憶装置110に記憶する。また、単語データベース部160は、発音矯正画像表示部242から制御部162を介して矯正情報コードが入力された場合に、入力された矯正情報コード(図18)に対応する矯正情報インデックス(図17(A)、(B))を記憶装置110

から読み出して、制御部162に対して出力する。

【0119】【比較部240】比較部240(図15)は、類似単語追加部186から入力される類似単語レコード(図18)が示す類似単語レコードに含まれる類似単語それぞれと、制御部162から入力される正解単語とを比較し、正解単語が類似単語のいずれかと一致するか否かを判断する。比較部240は、正解単語が類似単語と一致する場合にはその旨を、一致しない場合にはその旨を発音矯正画像表示部242に対して通知する。

【0120】【発音矯正画像表示部242】発音矯正画像表示部242は、比較部240が正解単語のいずれかと類似単語とが一致すると判定した場合に、類似単語レコード(図18)に付加された矯正情報コード(CCode)が示す矯正情報インデックス(図17(A),(B))が示す画像データおよびテキストデータの取得を制御部162に要求する。制御部162が、この要求に応じて画像データおよびテキストデータを記憶装置110から読み出し、発音矯正画像表示部242に対して出力すると、発音矯正画像表示部242は、これらのデータを、図19において符号(a)~(e)を付して例示したような位置に配置し、発音矯正画像を生成してモニタ102に表示する。

【0121】【音声識別・矯正プログラム22の動作】以下、音声識別・矯正プログラム22の動作を説明する。図20は、第2の実施形態における音声識別・矯正プログラム22(図15)の処理(S20)を示すフローチャートである。

【0122】図20において、ステップ200(S200)に示すように、ユーザの操作入力に応じて、制御部162が、例えば、図16に示したように、“read”という単語の発音をユーザに促す発音指示画像をモニタ102に表示する。制御部162は、正解単語“read”を比較部240に対して出力する。ユーザが発音指示画像に応じて“read”と発音すると、音声識別プログラム16のベクトルデータ作成部180、ラベルデータ作成部182、候補単語作成部184および類似単語追加部186(図15)は、ユーザが発音した音声进行を識別し、類似単語レコードを比較部240および発音矯正画像表示部242に対して出力する。

【0123】ここで、第1の実施形態において例示したように、ユーザが正しく“read”と発音すると、候補単語作成部184は単語“read”を含む候補単語を類似単語追加部186に対して出力し、類似単語追加部186は、単語“lead”等を類似単語として含む類似単語レコード(図18)を比較部240および発音矯正画像表示部242に対して出力する。反対に、例えば、ユーザが“r”の発音と“l”の発音とを区別できず、不正確な“read”の発音を行なうと、候補単語作成部184は、単語“read”の代わりに単語“lead”等を含む候補単語を類似単語追加部186に対して出力し、類似単語追加部186は、単語

語“read”等を類似単語として含む類似単語レコードを比較部240および発音矯正画像表示部242に対して出力する。

【0124】ステップ202(S202)において、比較部240は、制御部162から入力された正解単語と、類似単語追加部186から入力された類似単語レコードに含まれる類似単語それぞれとを比較し、正解単語が類似単語のいずれかと一致しない場合には、その旨を発音矯正画像表示部242に通知して“read”の発音の矯正・学習に関する処理を終了し、例えば次の単語の発音の矯正・学習に関する処理に進む。これ以外の場合には、比較部240は、正解単語が類似単語のいずれかと一致したことを発音矯正画像表示部242に通知して、S204の処理に進む。

【0125】ステップ204(S204)において、発音矯正画像表示部242は、類似単語レコードから矯正情報コード(CCode)を得る。さらに、発音矯正画像表示部242は、制御部162に対して矯正情報コードを出力し、発音矯正画像に用いる画像データおよびテキストデータの取得を要求する。制御部162は、発音矯正画像表示部242からの要求に応じて、単語データベース部160から、図17(A)に例示した矯正情報インデックスを得て、この矯正情報インデックスのエントリ(エントリ1~4,8)それぞれが示す画像データおよびテキストデータを記憶装置110から読み出し、発音矯正画像表示部242に対して出力する。

【0126】ステップ206(S206)において、発音矯正画像表示部242は、制御部162から入力された矯正情報インデックスのエントリ(エントリ1~4,8)それぞれに対応する画像データおよびテキストデータを、それぞれ図19に例示する位置(a)~(e)に配置した発音矯正画像を生成し、モニタ102に表示し、“read”に関する処理を終了し、例えば次の単語の発音の矯正・学習に関する処理に進む。

【0127】【変形例】なお、第2の実施形態においては、発音矯正情報がテキストデータおよび画像データのみを含む場合を例示したが、発音矯正情報が他の種類のデータを含んでいてもよい。例えば、発音矯正情報に正しい発音の音声データを含め、発音矯正情報(図19)をモニタ102に表示するとともに、音声データをスピーカ104を介して出力するようにしてもよい。

【0128】また、第2の実施形態においては、類似単語と発音矯正情報コードとを対応付けて管理する場合を示したが、図6に示した類似単語レコードに含めて管理するように音声識別・矯正プログラム22を構成してもよい。また、第2の実施形態においては、音声識別・矯正プログラム22が発音矯正情報のみを出力する場合を示したが、発音矯正情報および誤り情報の両方を出力するように構成してもよい。また、第2の実施形態においては、制御部162が発音指示画像をモニタ102に表

示し、発音矯正画像表示部242が発音矯正情報をモニタ102に表示する場合を例示したが、これらの構成部分のいずれかが、これら両方の画像をモニタ102に表示するように音声識別・矯正プログラム22を構成してもよい。

【0129】また、発音矯正画像(図19)に表示される全ての情報が矯正情報インデックス(図17(A)、(B))に登録されている必要はない。例えば、図17(A)に示した矯正すべき点および再発音支持のテキストデータは、音声識別・矯正プログラム22(図15)の作り方に応じて、矯正情報インデックスに登録されていても、あるいは、矯正情報インデックスには登録されず、発音矯正画像に予め書き込まれていてもよい。

【0130】[第3実施形態]以下、第3の実施形態として、第2の実施形態として示した発音矯正方法に応用した発音学習方法を説明する。この発音学習方法は、図20に示した発音識別・矯正プログラム22の動作を改良し、図16および図19に示した発音指示画像および発音の基礎を示す発音矯正画像の他に、発展的な学習を指示する発音矯正画像(図21~図23)をさらに表示し、ユーザの発音学習の便宜を図ったものである。

【0131】図21~図23は、それぞれ第3の実施形態として示す発音学習方法において用いられる第2~第4の発音矯正画像を例示する図である。第3の実施形態において、発音識別・矯正プログラム22は、図16、19に示した発音矯正画像に加え、図21~図23に例示する発音矯正画像を表示する。

【0132】図21に示す第2の発音矯正画像は、例えば、発音識別・矯正プログラム22が、第1の発音矯正画像(図19)を見てユーザが“read”を正しく発音したと判断した場合に表示され、ユーザが正しく“r”、“l”を区別して発音しているかを確認するために用いられる。なお、第2の発音矯正画像は、ユーザが第1の発音矯正画像に応じてユーザが正しく“read”と発音するまでは表示されず、ユーザが第2の発音矯正画像に示された単語“write”、“raw”、“long”、“light”の全てを正しく発音できるようになるまで繰り返し表示される。また、第2の発音矯正画像は、第1の発音矯正画像において省略されていた矯正情報インデックス(図17(A))のエントリ5、6を含んでいる。

【0133】図22に示す第3の発音矯正画像は、例えば、発音識別・矯正プログラム22が、第2の発音矯正画像(図21)内の各単語をユーザが正しく発音したと判断した場合に表示され、ユーザが“r”、“l”を区別するためのさらに進んだ練習を行なうために用いられる。なお、第3の発音矯正画像は、ユーザが第2の発音矯正画像に示された各単語の全てをユーザが正しく発音するまでは表示されず、第3の発音矯正画像に示されたセンテンス“write letters”、“great trouble”の全てをユーザが正しく発音できるようになるまで繰り返し表示され

る。また、第2の発音矯正画像は、第1の発音矯正画像において省略されていた矯正情報インデックス(図17(A))のエントリ7を含んでいる。

【0134】図23に示す第4の発音矯正画像は、例えば、発音識別・矯正プログラム22が、第3の発音矯正画像(図22)内のセンテンスの全てをユーザが正しく発音したと判断した場合に表示され、ユーザが“r”、“l”を区別して発音できるようになったことを確認するために用いられる。なお、第4の発音矯正画像は、ユーザが第3の発音矯正画像に示されたセンテンスの全てをユーザが正しく発音するまでは表示されず、第4の発音矯正画像に示されたセンテンス“The river rose several feet and finally overflowed its banks.”をユーザが正しく発音できるようになるまで繰り返し表示される。

【0135】[音声識別・矯正プログラム22の動作]以下、第3の実施形態における音声識別・矯正プログラム22の動作を説明する。図24は、第3の実施形態における音声識別・矯正プログラム22(図15)の動作(S30)を示すフローチャートである。図25は、図24に示した学習項目のリストアップ処理(S300)において作成される学習項目リストを例示する図である。

【0136】図24に示すように、ステップ300(S300)において、音声識別・矯正プログラム22は、図20に示したように、正解単語と類似単語とを比較し、類似単語と一致する正解単語を求める。さらに、音声識別・矯正プログラム22は、求めた正解単語において、ユーザが苦手とする発音(例えば“r”、“th”)を決定し、例えば図25に例示するように、学習項目としてリストアップする。

【0137】ステップ302(S302)において、音声識別・矯正プログラム22は、S300の処理においてリストアップした学習項目がまだ学習されずに残っているか否かを判断する。音声識別・矯正プログラム22は、学習項目が残っている場合にはS304の処理に進み、これ以外の場合には処理を終了する。

【0138】ステップ304(S304)において、音声識別・矯正プログラム22は、S300の処理においてリストアップした学習項目のひとつを取り出す。ステップ306(S306)において、音声識別・矯正プログラム22は、第2の実施形態に示したように、発音指示画像(図16)および発音矯正画像(図19)を表示し、ユーザに発音を学習させる。

【0139】ステップ308(S308)において、音声識別・矯正プログラム22は、ユーザの音声を識別し、識別の結果として得られた単語が候補単語と一致する場合にはユーザが正しく発音したと判断してS310の処理に進み、これ以外の場合にはユーザが正しく発音しなかったと判断してS306の処理に戻る。

【0140】ステップ310(S310)において、音

声識別・矯正プログラム22は、学習に用いていない応用問題（第2～第4の発音矯正画像；図21～図23）があるか否かを判断する。応用問題が残っている場合には発音識別・矯正プログラム22はS312の処理に進み、これ以外の場合にはS302の処理に戻る。

【0141】ステップ312（S312）において、音声識別・矯正プログラム22は、第2～第4の発音矯正画像のいずれかを表示し、ユーザに発音を学習させる。ステップ314（S314）において、音声識別・矯正プログラム22は、ユーザが正しく発音できたか否かを判断し、ユーザの発音が正しい場合にはS316の処理に進み、これ以外の場合にはS312の処理に戻る。

【0142】ステップ316（S316）において、音声識別・矯正プログラム22は、ユーザの発音に、S300の処理においてリストアップされた学習項目以外の誤りがあるか否かを判断する。このような誤りがある場合には、音声識別・矯正プログラム22はS318の処理に進み、これ以外の場合にはS310の処理に戻る。

【0143】ステップ318（S318）において、音声識別・矯正プログラム22は、S316の処理において見つかったユーザの発音上の誤りを学習項目に加え、S310の処理に戻る。

【0144】

【発明の効果】以上説明したように、本発明にかかる音声識別装置およびその方法によれば、上述した従来技術の問題点に鑑みてなされたものであり、所定の言語を母国語としない話者（ノン・ネイティブ）による所定の言語の話し声に含まれる単語それぞれを識別し、話者が意図する所定の言語の単語に置換して、正確なテキストデータを作成することができる。

【0145】また、本発明にかかる音声識別装置およびその方法によれば、話されている地域が異なる等のため、同一の言語の発音が変化した場合であっても、いずれの地域の話者による話し声でも、話者が意図する単語に変換して、正確なテキストデータを作成することができる。また、本発明にかかる音声識別装置およびその方法によれば、発音の個人差を補って、常に高い識別率を保つことができる。

【0146】さらに、本発明にかかる発音矯正装置およびその方法によれば、本発明にかかる上記音声識別装置およびその方法の処理の過程で得られるデータを利用して話者の発音の問題点を指摘することができ、また、話者にネイティブスピーカの発音を学習させ、話者の発音を矯正することができる。また、本発明にかかる発音矯正装置およびその方法によれば、話者の発音と正確な発音とを自動的に比較して誤りを指摘することができ、さらに、話者がどのように発音を矯正すべきかを示す詳細な情報を提示し、その発音を矯正することができる。

【図面の簡単な説明】

【図1】本発明にかかる音声識別処理を実現するコンピ

ュータの構成を例示する図である。

【図2】図2は、本発明にかかる音声識別処理を実現するソフトウェアの構成を示す図である。

【図3】図2に示した音声識別プログラムの構成を示す図である。

【図4】単語データのインデックステーブルに含まれるデータを例示する図である。

【図5】単語データの単語レコードに含まれるデータを例示する図である。

10 【図6】単語データの類似単語レコードに含まれるデータを例示する図である。

【図7】単語データの誤り情報コードテーブルを例示する図である。

【図8】候補単語作成部が類似単語追加部に出力する入力レコード(InWord)のデータ構造を示す図である。

【図9】候補単語作成部が類似単語追加部に出力する入力レコードマトリクス(InMatrix)のデータ構造を示す図である。

【図10】類似単語追加部が絞り込み部に出力する出力レコード(OutWord)のデータ構造を示す図である。

【図11】類似単語追加部が絞り込み部に出力する出力レコードマトリクス(OutMatrix)のデータ構造を示す図である。

【図12】コンピュータにおける本発明にかかる音声識別処理を示すフローチャート図である。

【図13】図13は、本発明にかかる音声識別処理および発音矯正方法を実現するコンピュータの構成を示す図である。

30 【図14】図14は、本発明にかかる音声識別処理および発音矯正方法を実現するソフトウェアを示す図である。

【図15】図14に示した音声識別・矯正プログラムの構成を示す図である。

【図16】図16は、図15に示した制御部が表示する発音指示画像を例示する図である。

【図17】図17（A）、（B）はそれぞれ、図15に示した単語データベース部が生成する矯正情報インデックスを例示する図であって、（A）はrの発音を矯正するための発音矯正画像を示し、（B）はlの発音を矯正するための発音矯正画像を示す。

【図18】図18は、第2の実施形態において単語データベース部が生成する類似単語レコードを示す図である。

【図19】図17（A）に例示した矯正情報インデックスが示す第1の発音矯正画像を例示する図である。

【図20】第2の実施形態における音声識別・矯正プログラム（図15）の処理（S20）を示すフローチャートである。

50 【図21】第3の実施形態として示す発音学習方法において用いられる第2の発音矯正画像を例示する図であ

る。

【図22】第3の実施形態として示す発音学習方法において用いられる第3の発音矯正画像を例示する図である。

【図23】第3の実施形態として示す発音学習方法において用いられる第4の発音矯正画像を例示する図である。

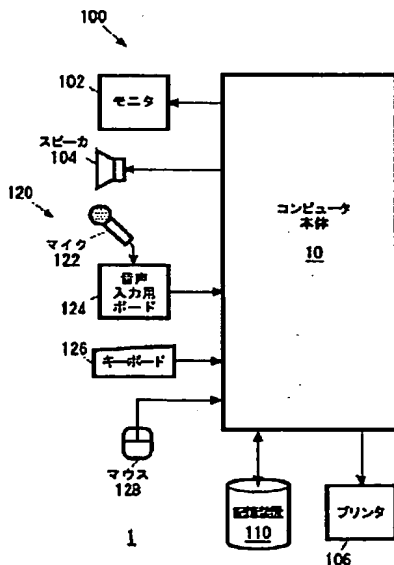
【図24】第3の実施形態における音声識別・矯正プログラム（図15）の動作（S30）を示すフローチャートである。

【図25】図24に示した学習項目のリストアップ処理（S300）において作成される学習項目リストを例示する図である。

【符号の説明】

- 1, 2・・・コンピュータ
10・・・コンピュータ本体
100・・・出力装置
102・・・モニタ
104・・・スピーカ
106・・・プリンタ
120, 130・・・入力装置
122・・・マイク
124・・・音声入力用ボード
126・・・キーボード

【図1】

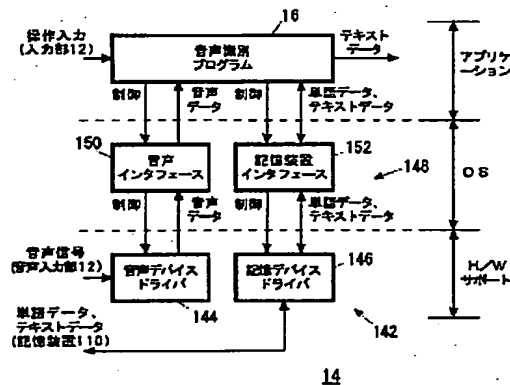


【図16】

"read"と発音してみてください！

- 128・・・マウス
132・・・画像入力用ボード
110・・・記憶装置
14, 20・・・ソフトウェア
142・・・ハードウェアサポート部
144・・・音声デバイスドライバ
146・・・記憶デバイスドライバ
148・・・オペレーティングシステム
150・・・音声インターフェース部
152・・・記憶装置インターフェース部
16・・・音声識別プログラム
160・・・単語データベース部
162・・・制御部
18・・・音声識別部
180・・・ベクトルデータ生成部
182・・・ラベル作成部
184・・・候補単語作成部
186・・・類似単語追加部
188・・・絞り込み部
22・・・音声識別・矯正プログラム
24・・・発音矯正プログラム
240・・・比較部
242・・・発音矯正画像表示部

【図2】



【図4】

インデックステーブル

Aへのポインタ (各アルファベットの先頭)	各アルファベット中の 単語レコードの数
⋮	⋮
Zへのポインタ	各アルファベット中の 単語レコードの数

【図7】

誤り情報コードテーブル

Ecode	内 容
0	OK
1	rをlと発音
2	lをrと発音
⋮	⋮
	lhを8と発音
⋮	⋮

【図8】

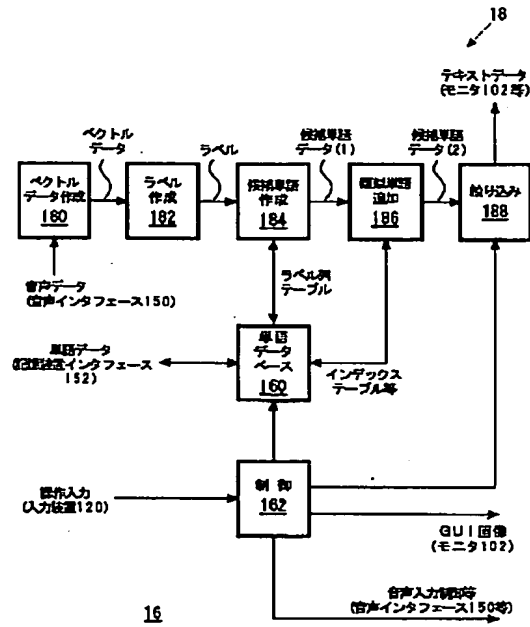
入力レコード InWord(i,j)

i	j	InWord
---	---	--------

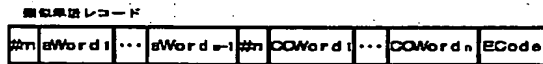
i: センテンス中 i 番目の単語
j: j 番目の辞書

InWord: 入力単語情報
- 単語
- 単語長
- 単語データ

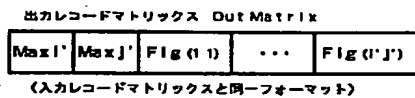
【図3】



【図6】



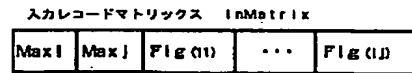
【図11】



【図5】

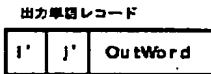


【図9】



Max I: センテンス中最大単語数
 Max J: 同一読みに対する最大候補数
 Fig (I J): I番目の単語にJ番目の候補があるか否かを示すフラグ

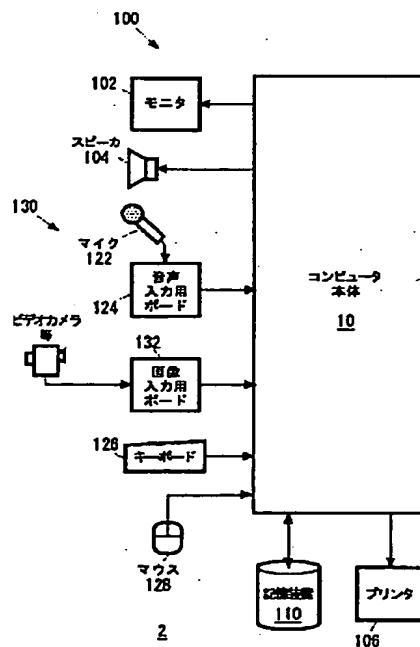
【図10】



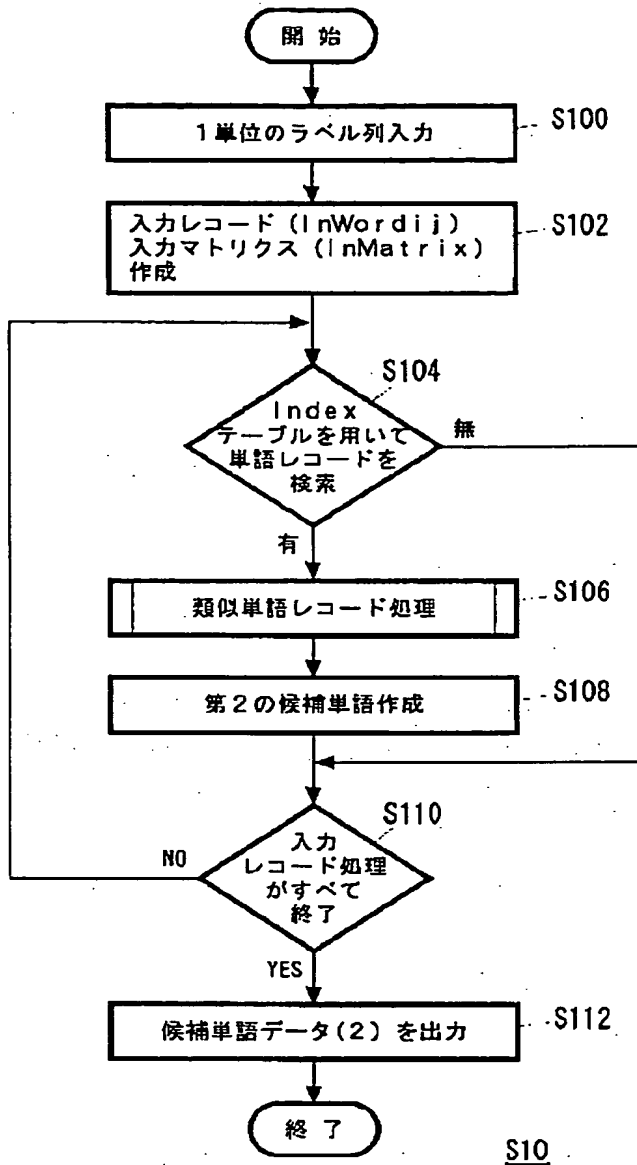
I': センテンス中I番目の単語
 J': I番目の候補

OutWord: 出力単語情報
 - 単語
 - 単語長
 - 属性データ
 - 入力単語との関連情報 (InWordに加えられる情報)

【図13】



【図12】



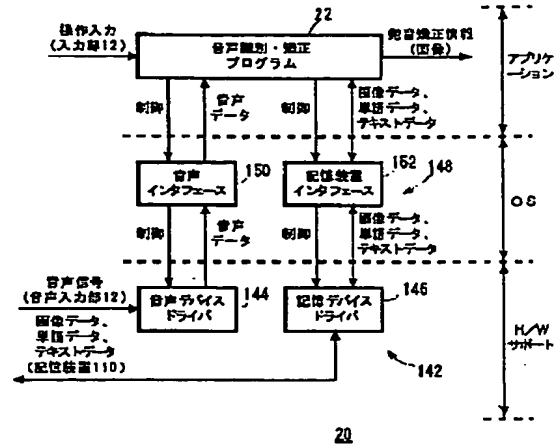
【図18】

類似単語レコード									
##	sWord1	...	sWord <i>i</i> -1	##	oWord1	...	oWord <i>n</i>	CCode	

【図25】

3	"r" (read)	"^" (hut)	"θ" (thin)
↑	↑		
エントリ数	(発音と例単語)		

【図14】



【図17】

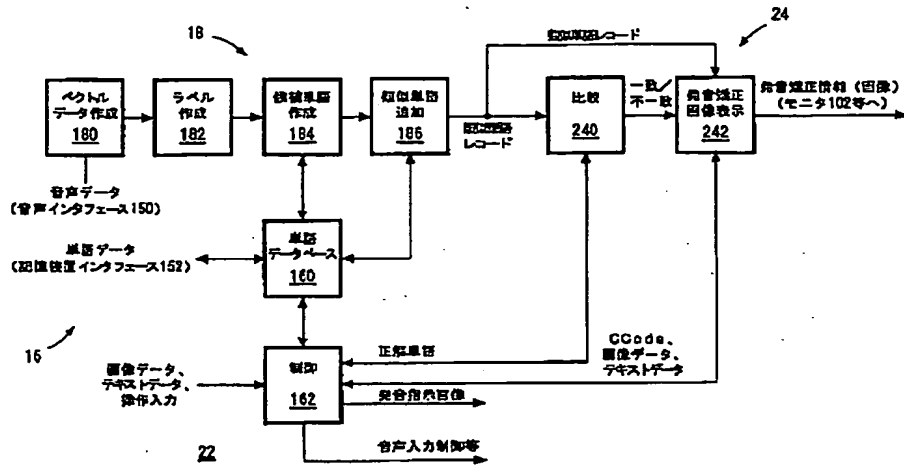
(A) CCode 1		
エントリ1	修正すべき点 (Text)	図19(a)
エントリ2	rの口の形 (Image)	図19(b)
エントリ3	lの口の形 (Image)	図19(c)
エントリ4	rと発音するためのアドバイス (Text)	図19(d)
エントリ5	rの単語例 (Text)	
エントリ6	lの単語例 (Text)	
エントリ7	rとlが含まれる単語の組み合わせ (Text)	図19(e) (図21、図22)
エントリ8	rとlが含まれる文章の例 (Text)	
	再発音指示 (Text)	図19(e) (図23)
(B) CCode 2		
エントリ1	修正すべき点 (Text)	
エントリ2	lの口の形 (Image)	
エントリ3	rの口の形 (Image)	
エントリ4	lと発音するためのアドバイス (Text)	
エントリ5	lの単語例 (Text)	
エントリ6	rの単語例 (Text)	
エントリ7	rとlが含まれる単語の組み合わせ (Text)	
エントリ8	rとlが含まれる文章の例 (Text)	
	再発音指示 (Text)	

【図21】

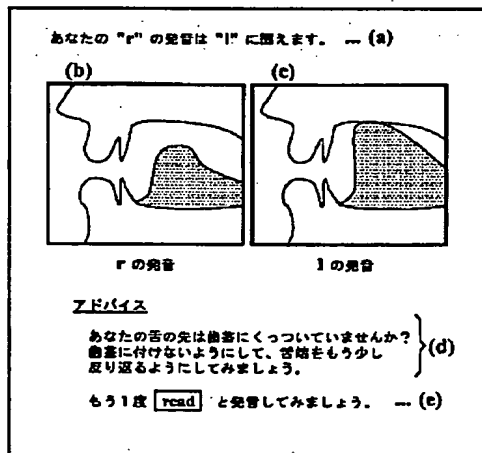
今度は **read** が正しく発音できました。
 それでは、本当に "r" の発音が理解できたか
 以下の単語を発音してみてください。

write **raw** (エントリ5)
long **light** (エントリ6)

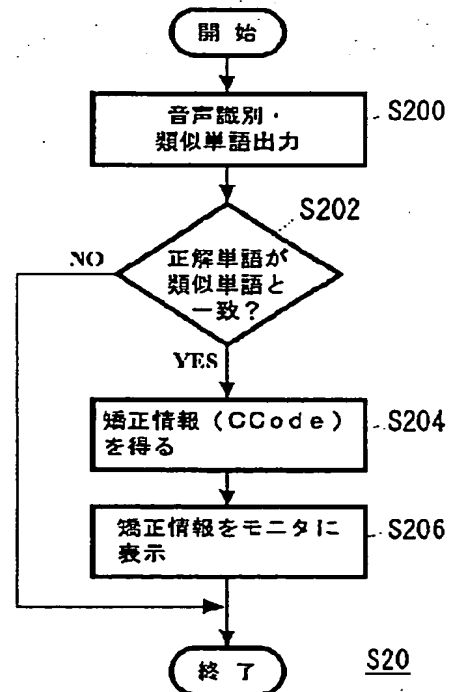
【図15】



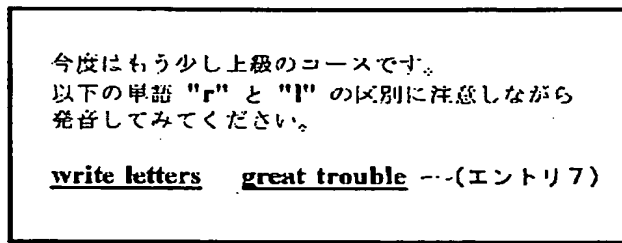
【図19】



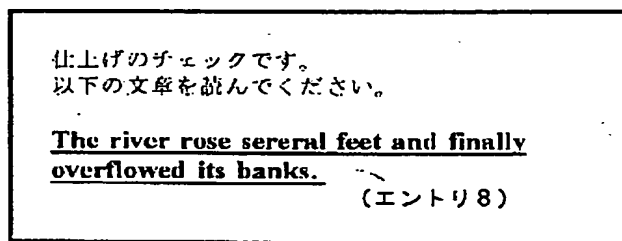
【図20】



【図22】



【図23】



【図24】

